Journal of the Korean Data & Information Science Society
2010, **21**(1), 51–59

한국데이터정보과학회지

가능도함수를 이용한 불연속점 수의 추정†

허집[1]

덕성여자대학교 정보통계학과

접수 2009년 11월 25일, 수정 2010년 1월 4일, 게재확정 2010년 1월 8일

요 약

일반화선형모형에서 회귀함수가 하나의 불연속점을 가질 때, Huh (2009)는 하나의 모수를 가지는 지수족의 가능도함수를 한쪽방향커널을 이용하여 그 불연속점의 위치와 점프크기를 추정하였다. 이 논문에서는 미지의 불연속점 수 q개를 가지는 회귀함수인 경우에, Huh (2009)가 제안한 점프크기 추정량의 점근분포를 이용한 가설검정법을 소개하고, 그 가설검정법을 이용한 불연속점 수를 추정하는 알고리듬을 제안하고, 모의실험을 통하여 추정의 정도를 알아보고자 한다.

주요용어: 국소다항적합, 로그가능도함수, 한쪽방향커널함수.

1. 서론

표본 $\{(X_i, Y_i); i = 1, 2, \cdots, n\}$은 확률벡터 (X, Y)로부터 독립적으로 추출되었다고 가정하자. 공변량 (covariate) X_i들은 토대 (support)가 $[0, 1]$인 확률밀도함수 $f(x)$를 가지며, $X = x$일 때 반응변수 Y의 조건부확률밀도함수는 아래

$$f_{Y|X}(y|x) = \exp\{y\theta(x) - b(\theta(x)) + c(y)\} \tag{1.1}$$

와 같이 하나의 모수를 가지는 지수족 (one parameter exponential family)이라 가정하자. 여기서 b와 c는 알려져 있는 함수이다.관심의 대상인 회귀함수/평균함수

$$m(x) \equiv E(Y|X = x) = b'(\theta(x))$$

의 일반화선형모형에서 모수적 추정은 연결함수 (link function) g를 이용하여 다음의 함수

$$\eta(x) = g(m(x))$$

를 선형으로 모형화하여 제안되고 있다. 만약 $g = (b')^{-1}$ 이면, 이때의 g를 정준 (canonical)연결함수라 한다. 자세한 내용은 McCullagh와 Nelder (1989)를 참조하기 바란다. 회귀함수 η의 비모수적 추정에 대해서는 Green과 Silverman (1994), Fan 등 (1995), Carroll 등 (1997), Huh와 Park (2002)등이 연구하였다.

회귀함수의 비모수적 추정에서 회귀함수가 불연속점 (discontinuity)을 가질 때 이러한 불연속점을 고려하지 않고 추정하는 경우 불연속점 주위에서 편의 (bias)를 가지게 되어 추정된 회귀함수가 일치

† 본 연구는 2008학년도 덕성여자대학교 연구비지원으로 이루어졌음.
[1] (132-714) 서울특별시 도봉구 쌍문동 419번지, 덕성여자대학교 정보통계학과, 부교수.
E-mail: jhuh@duksung.ac.kr

추정량 (consistent estimator)이 되지 않으며 실제구현에서도 문제점을 가지게 된다는 것을 Müller (1992)와 Huh와 Park (2004)이 선형모형에서 회귀함수가 불연속점을 가지는 연구에서 언급하였다.

Huh (2004, 2009)는 (1.1)과 같은 하나의 모수를 가지는 지수족의 회귀모형에서 회귀함수 m과 η의 ν차 미분함수 $\eta^{(\nu)}$가 한 점에서 불연속일 때 국소다항적합 (local polynomial fit)을 이용하여 불연속점의 위치 (location)와 점프크기 (jump size)를 커널가중국소로그가능도함수 (kernel weighted local log-likelihood function)를 바탕으로 제시하고 추정량들의 점근성질들을 규명하였다.

이 논문에서는 Huh (2009)가 제안한 불연속점의 점프크기 추정량을 이용하여 회귀함수 η의 불연속점의 존재여부에 대한 가설검정법을 제안하고자 한다. 또한, 회귀함수가 다중 (multiple) 불연속점을 가지고 그 수를 알지 못할 때, 그 수와 위치 및 점프크기를 추정하는 알고리듬을 소개하고자 한다.

선형모형에서 회귀함수의 불연속점의 존재여부에 대한 가설검정법과 그 수에 대한 비모수적 추정은 Yin (1988)과 Kim 등 (2003)에 의해 제안되었고, 분산함수의 불연속점의 존재여부의 가설검정법과 그 수에 대한 비모수적 추정은 Huh (2005, 2007, 2009)에 의해 연구되어졌다.

2절에서는 Huh (2009)가 제안한 회귀함수 η의 불연속점의 추정과 점근성질을 소개하고, 이를 이용하여 가설검정법을 제안한다. 3절에서는 2절에서 소개한 가설검정법으로 회귀함수가 가지는 불연속점의 수를 추정하는 알고리듬을 소개한다. 모의실험을 통하여 4절에서는 3절에서 제안한 알고리듬의 타당성을 소개하고자 한다.

2. 불연속에 대한 가설검정법

어떤 함수 a에 대하여 임의의 점 x를 중심으로 오른쪽과 왼쪽의 함수 극한값을 각각 $a_+(x) = \lim_{y\to x+} a(y)$와 $a_-(x) = \lim_{y\to x-} a(y)$로 정의하자. 회귀함수 η가 한 점 τ에서 불연속인 경우 점프크기는

$$\Delta = \eta_+(\tau) - \eta_-(\tau) \tag{2.1}$$

로 표현된다. 불연속점이 존재한다면 식 (2.1)에서 $|\Delta| > 0$이고, 그렇지 않다면 $\Delta = 0$이다.

식 (1.1)의 조건부확률밀도함수 $f_{Y|X}(y|x)$를 회귀함수 η로 표현하면 다음과 같이

$$f_{Y|X}(y|x) = \exp\left\{y(g\circ b')^{-1}(\eta(x)) - b((g\circ b')^{-1}(\eta(x))) + c(y)\right\} \tag{2.2}$$

표현된다. 여기서 $\circ$는 합성함수를 나타낸다. 식 (2.2)의 로그변환 함수를 $\ell(\eta(x), y)$로 표현하자.

회귀함수 η를 $p(\geq\ 0)$차 국소다항식으로 근사적으로 표현하고, 로그 변환된 조건부확률밀도함수 $\ell(\eta(x), y)$를 이용하면 다음과 같은 커널가중국소로그가능도함수

$$\sum_{i=1}^{n} \ell(\sum_{j=1}^{p} \alpha_j (X_i - x)^j, Y_i) K\left(\frac{X_i - x}{h}\right) \tag{2.3}$$

를 생각할 수 있다. 여기서 함수 K는 토대가 $[0,1]$인 한쪽방향커널함수 (one-sided kernel function)이며 h는 띠폭 (bandwidth)이다. 함수 $\ell(u, y)$는 정준연결함수를 선택하게 되면 u에 대하여 볼록함수 (convex function)이므로 식 (2.3)을 최대로 하는 해가 존재하며 한쪽방향커널함수는 x를 기준으로 오른쪽의 표본만 가중치를 주는 역할을 하게 된다. 따라서 식 (2.3)을 최대로 하는 해를 $\hat{\underset{\sim}{\alpha}}^+ = (\hat{\alpha}_0^+, \hat{\alpha}_1^+, \cdots, \hat{\alpha}_p^+)^T$라 두면 $\eta_+(x)$의 추정량으로 $\hat{\eta}_+(x) = \hat{\alpha}_0^+$라 제안할 수 있다. 이와 같이 식 (2.3)에서 $K(h^-1(X_i - x))$대신 $K(h^-1(x - X_i))$를 사용하면 x를 기준으로 왼쪽의 표본만 가중치를 주게 되어 식 (2.3)을 최대로 하는 해를 $\hat{\underset{\sim}{\alpha}}^- = (\hat{\alpha}_0^-, \hat{\alpha}_1^-, \cdots, \hat{\alpha}_p^-)^T$라 두면 $\eta_-(x)$의 추정량으로 $\hat{\eta}_-(x) = \hat{\alpha}_0^-$라 제안할 수 있다.

임의의 점 x에서 점프크기의 추정량으로 다음

$$\widehat{\Delta}(x) = \widehat{\eta}_+(x) - \widehat{\eta}_-(x) \tag{2.4}$$

와 같이 정의하면 회귀함수 η가 하나의 불연속점 τ를 가지는 경우, τ의 추정량은 다음

$$\widehat{\tau} = \inf\left\{ z \in Q : |\widehat{\Delta}(z)| = \sup_{x\in Q}|\widehat{\Delta}(x)| \right\} \tag{2.5}$$

으로 제안할 수 있다. 여기서 Q는 $Q \subset (0,1)$인 폐구간이다. 또한 불연속점 τ에서 점프크기 Δ의 추정량은 자연스럽게 다음의 식

$$\widehat{\Delta}(\widehat{\tau}) = \widehat{\eta}_+(\widehat{\tau}) - \widehat{\eta}_-(\widehat{\tau}) \tag{2.6}$$

으로 정의할 수 있다.

임의의 $x \in Q$에서 η가 불연속인지에 대한 귀무가설과 대립가설은

$$H_0 : \Delta(x) = 0, H_1 : \Delta(x) \neq 0$$

과 같이 표현할 수 있고 위 가설에 대한 가설검정을 위하여 점프크기 추정량 (2.6)을 검정통계량으로 활용할 수 있다. Huh (2009)는 커널함수 K와 띠폭 h의 적절한 조건하에서 식 (2.5)의 불연속점의 위치 추정량 $\widehat{\tau}$의 수렴속도가 n^{-1}이 됨을 보였고, 식 (2.6)의 점프크기 추정량 $\widehat{\Delta}(\widehat{\tau})$의 점근분포가 다음

$$\frac{1}{f(\tau)}\left(\frac{1}{v_+(\tau)} + \frac{1}{v_{-(\tau)}}\right)\int_0^1 (K(u)^2 du) \tag{2.7}$$

과 같음을 보였다. 여기서 $v(x) = Var(Y|X = x)$는 x에서의 분산함수 (variance function)이다.

Huh (2009)는 위 결과에서 점근분포의 분산에 포함되어 있는 장애모수 (nuisance parameter)인 점근분포의 분산에 내재해 있는 v와 f의 추정량을 제시하면 불연속점의 존재 유무에 대한 가설검정을 할 수 있다고 설명하였다. 본 연구에서는 v와 f의 추정량으로 다음과 같이

$$\widehat{f}(x) = \frac{1}{nh_1}\sum_{i=1}^{n} L\left(\frac{X_i - x}{h_1}\right), \tag{2.8}$$

$$\widehat{v}(x) = \frac{\frac{1}{nh_1}\sum_{i=1}^{n} L\left(\frac{X_i - x}{h_1}\right)\{Y_i - \widehat{m}(X_i)\}^2}{\frac{1}{nh_1}\sum_{i=1}^{n} L\left(\frac{X_i - x}{h_1}\right)} \tag{2.9}$$

커널형 추정량으로 제시하자. 여기서 $\widehat{m}$은 회귀함수 m의 Nadaraya-Watson 추정량이며 L은 대칭인 확률밀도함수로 토대가 $[-1,1]$을 가지는 커널함수이고 h_1는 띠폭이다. 식 (2.8)는 커널형 확률밀도함수이며 식 (2.9)는 분산함수의 Nadaraya-Watson 커널추정량이다. Stute (1982)에 의해 $\widehat{f}(x)$는 x에 의해 균일일치성 (uniform consistency)이 규명되어졌고, $\widehat{v}(x)$ 도 귀무가설 $H_0 : \Delta = 0$ 하에서 적절한 조건을 만족하는 경우에 균일일치성을 만족함을 Huh (2006)의 Theorem 1의 증명에 따라서 쉽게 보일 수 있다.

위에서 언급한 확률밀도함수와 분산함수의 균일일치성과 (2.7)의 결과에 의하여 귀무가설 $H_0 : \Delta = 0$ 하에서 분산함수는 연속이며 $v_+(\tau) = v_-(\tau)$이므로, 다음

$$\sqrt{nh}\frac{\sqrt{\widehat{f}(\widehat{\tau})}\widehat{\Delta}(\widehat{\tau})}{\sqrt{\frac{2}{\widehat{v}(\widehat{\tau})}\int_0^1 (K(u))^2 du}} \xrightarrow{d} N(0,1) \tag{2.10}$$

이 성립함을 알 수 있다. 따라서, 식 (2.10)을 이용하여 귀무가설 $H_0 : \Delta = 0$의 가설검정법을 제안할 수 있다. 불연속점의 점프크기 추정량의 점근분포의 결과를 이용하여 위와 같이 불연속점의 존재 유무에 대한 가설검정법을 Huh (2006, 2009)는 분산함수에 적용하여 연구하였다.

3. 불연속점 수의 추정

Kim 등 (2007)은 2절에서 점프크기 추정량의 점근분포를 이용하여 제안한 가설검정법을 각각 선형모형의 회귀함수와 분산함수의 불연속점 수에 대한 추정으로 제안하였다. 이 절에서는 2절에서 불연속점의 존재유무에 대한 가설검정법으로 일반화선형모형의 회귀함수 η가 가지는 미지의 불연속점의 수를 추정하는 알고리듬을 제안한다.

회귀함수 η는 알려져 있지 않은 불연속점의 수 q개를 가지며 그 불연속점의 위치는 각각 $\tau_j \in (0,1)$, $j = 1, \cdots, q$라고 가정하자. 그리고 $\Delta(\tau_j)$는 각각 τ_j에서 불연속점의 점프크기이며, η는 다음의 식

$$\eta(x) = \xi(x) + \sum_{j=1}^{q} \Delta(\tau_j) \times \mathrm{I}[x \geq \tau_j]$$

으로 표현된다고 하자. 여기서 I는 지시함수 (indicator function)이다. 불연속점들 τ_j, $j = 1, \cdots, q$에서 각 불연속점들의 점프크기 $\Delta(\tau_j)$는 다음의 식

$$\Delta(\tau_j) = \eta_+(\tau_j) - \eta_-(\tau_j)$$

으로 표현된다.

설명의 편의를 위하여 점프크기 $\Delta(\tau_j), j = 1, \cdots, q,$ 들이 다음의 조건

$$|\Delta(\tau_1)| \ \geq \ |\Delta(\tau_2)| \ \geq \cdots \geq \ |\Delta(\tau_q)|$$

을 만족한다고 가정하자. 직관적인 관점에서, 점프크기 추정값 $\widehat{\Delta}(x)$의 절대값이 가장 크다면, 점 x에서 가장 큰 점프크기를 가지는 불연속점의 위치라고 추정할 수 있다. 즉, 집합 $Q_1 = \{x : \ h \leq x_k \leq 1 - h\}$에서 가장 큰 점프크기를 가지는 불연속의 위치 τ_1의 추정량 $\widehat{\tau}_1$은 다음과 같이

$$\widehat{\tau}_1 = \arg\max_{Q_1} \left|\widehat{\Delta}(x)\right|$$

으로 정의할 수 있다. 여기서 $Q_1 = [h, 1-h]$으로 선택한 것은 토대 $[0,1]$에서 0과 1은 또 다른 형태의 불연속점으로 이해될 수 있고, 이러한 이유로 여러 연구 논문 (Müller, 1992; Loader, 1996; Jose와 Ismail, 1999; Huh와 Carrire, 2002; Huh와 Park, 2004)에서 불연속점의 존재 구간으로 $[h, 1-h]$가 선택되어졌기 때문이다. 위치추정량 $\widehat{\tau}_1$에서 불연속점이 있는지를 알아보기 위해 다음의 가설

$$H_0 : \Delta(\tau_1) = 0, \quad H_1 : \Delta(\tau_1) \neq 0 \tag{3.1}$$

을 생각하고 $\widehat{\tau}_1$에서 점프크기 추정량 $\widehat{\Delta}_1(\widehat{\tau}_1)$의 귀무가설 하에서 점근분포 (2.10)을 이용하여 위 가설에 대한 검정통계량으로 다음과 같이

$$T(\widehat{\tau}_1) = \sqrt{nh}\frac{\sqrt{f(\widehat{\tau}_1)}\widehat{\Delta}(\widehat{\tau}_1)}{\sqrt{\frac{2}{\widehat{v}(\widehat{\tau}_1)}\int_0^1 (K(u))^2 du}} \xrightarrow{d} N(0,1) \tag{3.2}$$

제시할 수 있다. 식 (2.10)에 의하여 귀무가설 $\Delta(\tau_1) = 0$하에서 $T(\widehat{\tau}_1) \xrightarrow{d} N(0,1)$이다. 그러므로 가설 (3.3)에 대한 기각역은 유의수준 α에서 다음과 같이

$$|T(\widehat{\tau}_1)| > z_{\alpha/2} \tag{3.3}$$

로 주어진다. 여기서 z_α는 표준정규분포의 (1- α)분위수이다.

두 번째로 큰 점프크기를 가지는 불연속점의 위치 τ_2의 추정량은 다음과 같이

$$\widehat{\tau}_2 = \arg\max_{Q_2} \left|\widehat{\Delta}(x)\right|$$

정의할 수 있다. 여기서 $Q_2 = \{x : h \le x \le 1-h, \ \ |x - \widehat{\tau}_1| \ge 2h\}$이다. 한 불연속점의 점프크기 추정치들이 인접해 있는 x의 점프크기에 영향을 주기 때문에 하나의 불연속점 주변의 점들에서도 불연속점으로 판정될 가능성이 있다. 따라서, Jose와 Ismail (1999)이 언급하였던 것처럼 근접해 있는 두 개의 불연속점들 사이의 거리들은 $2h$ 보다 크다는 가정이 필요하게 된다. 따라서 Q_2는 $\widehat{\tau}_1$ 으로부터 $2h$ 떨어진 점들로 구성된 집합이다. 두 번째 불연속점에 대한 가설 $H_0 : \Delta(\tau_2) = 0, \ \ H_1 : \Delta(\tau_2) \ne 0$을 생각하면 검정통계량 (3.4)의 $T(\widehat{\tau}_2)$를 이용하여 (3.5)와 같이 기각역 $|T(\widehat{\tau}_2)| > z_{\alpha/2}$이 주어진다.

이와 같은 방법으로 다음으로 큰 점프크기를 가지는 τ_3, τ_4 등의 추정량으로 $\widehat{\tau}_3$, $\widehat{\tau}_4$ 등을 순차적으로 다음과 같이

$$\widehat{\tau}_3 = \arg\max_{Q_3} \left|\widehat{\Delta}(x)\right|,$$
$$\widehat{\tau}_4 = \arg\max_{Q_4} \left|\widehat{\Delta}(x)\right|,$$

제시할 수 있다. 여기서 $Q_j = \{x : h \le x \le 1-h, \ \ |x - \widehat{\tau}_i| \ge 2h, \ i = 1, \cdots, j-1\}$, $j = 1, 2, \ldots$ 는 j번째로 점프크기가 큰 불연속점의 존재영역에 대한 구간이다. 단계 j에서의 가설 $H_0 : \Delta(\tau_j) = 0, \ \ H_1 : \Delta(\tau_j) \ne 0$에 대한 기각역은 다음과 같이 $|T(\widehat{\tau}_j)| > z_{\alpha/2}$주어진다. 따라서 $\widehat{\Delta}(\widehat{\tau}_j)$의 크기 순으로 $\widehat{\tau}_j$에 대해 순차적으로 가설검정을 해 나가면서, 귀무가설이 기각되지 않은 단계가 $j = r$이라면 미지의 불연속점 수 q의 추정량을 $\widehat{q} = r - 1$로 제안할 수 있다.

4. 모의실험

이 절에서는 3절에서 소개한 일반화선형모형의 회귀함수의 불연속점 수의 추정에 대한 알고리듬을 모의실험을 통하여 구현해 보고자 한다. 모의실험을 위하여 설명변수인 X_i의 확률밀도함수가 [0, 1]구간에서의 균등분포인 경우를 생각하고, 반응변수 Y_i의 분포는 베르누이분포를 고려해 보자. m은 아래와

$$m(x) = \frac{\exp(\eta(x))}{1 + \exp(\eta(x))}$$

같이 선택하였고, 여기서

$$\eta(x) = \begin{cases} -20x, & 0 \le x \le 0.3 \\ -20(x-1), & 0.3 < x \le 0.75 \\ 20(x-1), & 0.75 < x \le 1 \end{cases}$$

이다. 회귀함수 η은 두 개의 불연속점 $\tau_1 = 0.3$ 과 $\tau_2 = 0.75$를 가진다. 각 불연속점의 점프크기는 $\Delta_1 = 20$ 과 $\Delta_2 = -10$이다. 정준연결함수인 로짓연결함수 (logit link function) $g(u) = \log(u/(1-u))$를 선택하였고, 국소다항적합에서의 다항식의 차수는 $p = 0$로 하였다.식 (2.3)의 커널가중국소로그가능도함수에 사용된 한쪽방향커널함수는 biweight 커널함수를 이용하여 만든 것으로 다음의 커널을

$$K(x) = \frac{15}{8}(1-x^2)^2 \times \mathrm{I}[0 \le x \le 1]$$

선택하였다. 또한 (2.8)과 (2.9)의 검정통계량에 내에 있는 추정된 불연속점에서의 커널형 확률밀도함수, 회귀함수와 분산함수의 추정량에 이용된 커널은 biweight 커널로서 다음과 같이

$$L(x) = \frac{15}{16}(1-x^2)^2 \times \mathrm{I}[-1 \le x \le 1]$$

를 선택하였다.

불연속점의 추정을 위해서 띠폭 h는 0.15를 선택하였고, 식 (2.8)과 (2.9)의 장애모수의 추정을 위한 띠폭 h_1도 0.15로 선택하여 모의실험 결과를 연구하였다. 3절에서 설명하였듯이, 점프크기가 가장 큰 불연속점의 위치를 추정하기 위하여 구간 Q_1은 $[h, 1-h]$로 선택하였고, 구간 $[0,1]$을 100등분한 $t_k = k/100$, $k = 1, \cdots, 100$에서 각각의 점프크기 추정치를 구하였다. 제안된 불연속점 수의 추정을 위한 각 단계별 가설검정법에서 유의수준은 $\alpha = 0.05$로 하였다.

불연속점 수와 그 위치 추정의 정도를 살펴보기 위하여 표본의 수를 $n = 1,000$으로 하여 1,000회 반복하였다. 불연속점 수 q에 대한 추정 $\hat{q}$의 평균과 표준오차는 각각 2.083과 0.008724이다. 아래 표 4.1은 불연속점 수의 추정량 $\hat{q}$과 각각의 빈도를 제시하여 모의실험 결과에 의한 $\hat{q}$의 분포를 살펴보았다. 점프크기가 가장 큰 첫 번째 불연속점 위치 추정의 정보를 보기 위하여 표 4.2에서 τ_1이 추정된 위치 t_k의 지표 k와 해당되는 빈도를 나열하였다. 첫 번째 불연속점 $\tau_1 = 0.3$에 해당되는 $k = 30$인 경우 1,000번 중 868번이 불연속점으로 판정되었다. 위치 $k = 75$인 곳은 두 번째로 점프크기가 큰 $\tau_2 = 0.75$에 해당되는 지점이지만 1,000번 중 128번이 가장 큰 점프크기를 가진 불연속점의 위치로 판정하였다. 표 4.3은 두 번째로 큰 점프크기를 가지는 불연속점의 위치 $\tau_2 = 0.75$의 추정 정보를 보여주고 있다. $\tau_2 = 0.75$에 해당되는 $k = 75$인 경우에 1,000번 중 566번이 불연속점으로 판정되었고, 그 주변에서도 불연속점으로 판정되고 있다. 점프크기가 가장 큰 τ_1의 위치 $k = 30$을 두 번째로 점프크기가 큰 불연속점의 위치로 판정하고 있다. 두 추정량 $\hat{\tau}_1$과 $\hat{\tau}_2$이 편의를 보이고 있지만, 점프크기 순으로 순차적으로 각각의 참값인 0.3와 0.75를 잘 추정하고 있다. 한편 표 4.1에서 보여주듯이, 불연속점 수를 3으로 추정한 빈도가 1,000번 중 83번이었는데, 세 번째 불연속점으로 추정된 위치에 대한 결과를 표 4.4에서 보여주고 있다.

지금까지 일반화선형모형에서 회귀함수의 불연속점에 대한 연구가 없었기에 모의실험의 결과를 비교할 수 없어 본 연구에서 제안한 알고리듬의 추정 정도만 설명되었지만, 불연속점 수와 그 위치들의 추정의 정도는 좋다고 할 수 있다. Huh (2009)는 선형모형에서 Hart와 Yi (1998)가 띠폭의 선택으로 제안한 한쪽방향커널함수를 사용한 교차타당성 (cross-validation) 방법을 활용하여 불연속점의 띠폭 선택을 설명하였으나, 향후 일반화선형모형에서의 교차타당성을 이용하여 띠폭 선택 방법을 연구할 필요가 있

표 **4.1** 불연속점 수의 추정치 $\widehat{q}$의 빈도

$\widehat{q}$	2	3
빈도	917	83

표 **4.2** 첫 번째 불연속점의 위치추정치 $\widehat{\tau}_1$의 t_k에서의 빈도

k	26	27	28	29	30	75	계
빈도	1	1	1	1	868	128	1000

표 **4.3** 두 번째 불연속점의 위치추정치 $\widehat{\tau}_2$ 의 t_k에서의 빈도

k	30	42	51	53	57	58	59	60	61	62	63	64
빈도	127	1	1	1	1	2	2	1	2	7	9	9
k	65	66	67	68	69	70	71	72	73	74	75	계
빈도	13	8	13	20	16	32	31	36	39	63	566	1000

표 **4.4** 세 번째 불연속점의 위치추정치$\widehat{\tau}_3$의 t_k에서의 빈도

k	26	42	43	44	49	55	56	57	58	59	75
빈도	1	1	1	1	1	2	2	1	2	7	9
k	76	77	78	79	80	81	82	83	85	계	
빈도	13	8	13	20	16	32	31	36	39	83	

다. 또한, 기존의 불연속점 추정이 연구되어진 논문들과 같이 본 연구에서도 나타난 추정치의 편의에 대한 연구로 그 원인을 규명할 필요가 있다.

참고문헌

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the Amrican Statistical Association*, **92**, 477-489.

Fan, J., Heckman, N. E. and Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the Amrican Statistical Association*, **90**, 141-150.

Green P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*, Chapman and Hall, London.

Hart, J. D. and Yi, S. (1998). One-sided cross-validation. *Journal of the Amrican Statistical Association*, **93**, 620-631.

Huh, J. (2004). Nonparametric discontinuity point estimation in generalized linear model. *Journal of the Korean Statistical Society*, **33**, 59-78.

Huh, J. (2005). Nonparametric detection of a discontinuity point in the variance function with the second moment function. *Journal of Korean Data & Information Science Society*, **16**, 591-601.

Huh, J. (2006). Testing the existence of a discontinuity point in the variance function. *Journal of Korean Data & Information Science Society*, **17**, 707-716.

Huh, J. (2007). Nonparametric detection algorithm of discontinuity points in the variance function. *Journal of Korean Data & Information Science Society*, **18**, 669-678.

Huh, J. (2009). Detection of a change point based on local-likelihood. *Journal of Multivariate Analysis*, in revision.

Huh, J. (2009). Testing of a discontinuity point in the log-variance function based on likelihood. *Journal of Korean Data & Information Science Society*, **20**, 1-9.

Huh, J. and Carriére, K. C. (2002). Estimation of regression functions with a discontinuity in a derivative with local polynomial fits. *Statistics and Probability Letters*, **56**, 329-343.

Huh, J. and Park, B. U. (2002). Likelihood-based local polynomial fitting for single-index models. *Journal of Multivariate Analysis*, **80**, 302-321.

Huh, J. and Park, B. U. (2004). Detection of change point with local polynomial fits for random design case. *Australian and New Zealand Journal of Statistics*, **46**, 425-441.
Jose, C. T. and Ismail, B. (1999). Change points in nonparametric regression functions. *Communication in Statistics-Theory and Methods*, **28**, 1883-1902.
Kim, J. T., Choi, H. and Huh, J. (2003). Detection of change-points by local linear regression fit. *The Korean Communications in Statistics*, **10**, 31-38.
Loader, C. R. (1996). Change point estimation using nonparametric regression. *Annals of Statistics*, **24**, 1667-1678.
McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, 2nd ed., Chapman and Hall, London.
Müller, H. G. (1992). Change-points in nonparametric regression analysis. *Annals of Statistics*, **20**, 737-761.
Stute, W. (1982). A law of the logarithm for kernel density estimators. *Annals of Probability*, **10**, 414-422.
Yin, Q. (1988). Detection of the number, locations and magnitudes of jumps. *Communications in Statistics-Stochastic Models*, **4**, 445-455.

Journal of the Korean Data & Information Science Society
2010, **21**(1), 51–59

한국데이터정보과학회지

Estimation of the number of discontinuity points based on likelihood[†]

Jib Huh[1]

[1]Department of Statistics, Duksung Women's University

Received 25 November 2009, revised 4 January 2010, accepted 8 January 2010

Abstract

In the case that the regression function has a discontinuity point in generalized linear model, Huh (2009) estimated the location and jump size using the log-likelihood weighted the one-sided kernel function. In this paper, we consider estimation of the unknown number of the discontinuity points in the regression function. The proposed algorithm is based on testing of the existence of a discontinuity point coming from the asymptotic distribution of the estimated jump size described in Huh (2009). The finite sample performance is illustrated by simulated example.

Keywords: Local polynomial fit, log-likelihood function, one-sided kernel function.

[†] This research was supported by the Duksung Women's University Research Grants 2008.

[1] Associate Professor, Department of Statistics, Duksung Women's University, Seoul 132-714, Korea.
E-mail: jhuh@duksung.ac.kr